DOCUMENT RESUME

ED 228 916 HE 016 076

AUTHOR Boak, Terry; And Others
TITLE The Evaluation of Faculty Teaching Performance.
INSTITUTION Memorial Univ., St. John's (Newfoundland). Faculty of Education.
PUB DATE Jan 83
NOTE 100p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *College Faculty; College Instruction; Evaluation Criteria; Evaluation Methods; *Faculty Evaluation; Feedback; *Formative Evaluation; Higher Education; Personnel Policy; Research Problems; *Student Evaluation of Teacher Performance; *Summative Evaluation; *Teacher Effectiveness; Teaching Skills
IDENTIFIERS Memorial University of Newfoundland (Canada)

ABSTRACT

A faculty evaluation study conducted at Memorial University of Newfoundland for the Committee on Faculty Evaluation is described. Evaluation criteria/methods and recent research findings regarding student ratings of faculty instruction are considered, along with the advantages and disadvantages of formative and summative evaluations. Formative theory, which is based largely on constructive feedback, is considered in regard to: teacher characteristics, teacher effectiveness, teacher utility, and research issues (the problems of measuring change, nonlinearity, and measuring utility function constraints). The benefits of summative evaluation, which provides information for personnel decisions, are weighed against implementation problems, including selecting teaching dimensions to be included on rating scales, and its bureaucratic consequences. Recommendations include the following: the university should not use summative evaluation of teaching performance (i.e., no student, administrator, or peer evaluations of teaching performance). Appendices include a proposed compromise to the problem of evaluating teacher effectiveness and the following articles by J. W. Bulcock: "A Respecified Model of Teaching Effectiveness" and "Why Is the Matter of Faculty Evaluation a Pressing Issue in the 1980s?" (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED228916

THE EVALUATION OF FACULTY TEACHING PERFORMANCE

A Report

Prepared by the Committee on Student Evaluation

Terry Boak
Jeff Bulcock (Chairman)
Mark Glassman
Alan Griffiths

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jeff Bulcock

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

For presentation to the Committee on Faculty Evaluation

Roy Kelleher
Glen Sheppard
Phil Warren (Chairman)

HE 016 076

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

January 1983

REPORT SUMMARY

Title: The Evaluation of Faculty Teaching Performance

A Report Prepared by:

The Committee on Student Evaluation
Faculty of Education
Memorial University of Newfoundland
St. John's, Newfoundland
Canada A1B 3X8

Committee Members:

Dr. Terry Boak
Department of Educational Psychology

Mr. Jeff Bulcock, Chairman
Department of Educational Foundations

Dr. Marc Glassman
Department of Curriculum and Instruction

Dr. Alan Griffiths
Department of Curriculum and Instruction

Summary

Of the several reasons for evaluating the teaching performance of university faculty two are paramount. First, an evaluation provides the individual with feedback on his teaching which may be used to improve his teaching in the future. Second, an evaluation may be used to provide information which administrators and tenure committees may use to make personnel decisions relating to tenure, promotion and merit salary increments. The first kind is known as formative evaluation while the second is known as summative evaluation. The analysis supported the use of formative evaluation but not use of summative evaluation.

Most teachers can significantly improve their teaching effectiveness if they are given constructive feedback on their teaching and the opportunity and encouragement to develop a teaching style which is congruent with their personality characteristics. The formative model of teaching does not, however, support a single model of teaching; but rather a family of models. This is because teachers with different

personality characteristics and different perceptions of teaching develop different styles. The different styles may be equally effective though little research evidence on this question exists.

Formative models of teaching are not static. Thus, conventional cross-sectional research designs are inappropriate as tests of the propositions supported by the model. Instead, time-series models with lagged variables are necessary even though they are complex and difficult to estimate accurately. Unfortunately, dynamic models of teaching effectiveness designed to test a formative function of teaching performance have not been estimated.

The summative evaluation of teaching calls for a formal assessment of teaching performance. Teaching is usually coequal with research and scholarship when it comes to evaluating a faculty member's institutional utility. The terms and conditions of university employment often call for the evaluation of faculty; thus, summative evaluation in some form is here to stay. Many university faculty members and many faculty associations favour the summative evaluation of teaching - especially student evaluation - because it provides protection against arbitrary personnel decision-making.

In theory, then, there is a good deal of support for summative evaluation. In practice, however, there are several serious limitations to the procedure. Administrators tend not to like evaluating teaching. They are far more likely to support student or colleague evaluation. When they do evaluate they tend to inflate their assessments. Observers, whether administrators or students, bring their own set of assumptions to the teaching context which significantly influence their interpretation. In other words, the observer influences the process - a serious source of bias. In addition it has been found that both students and colleagues are easily fooled by apparently slick, enthusiastic and dynamic teachers who were "trained" to teach nonsense.

There are other reservations about summative evaluation as well. For example, the models of teaching effectiveness have not demonstrated the existence of a strong relationship between what is defined as effective teaching and student achievement. This is probably because most models of teacher effectiveness are seriously misspecified. Many models are composed of numerous dimensions of teaching which are thought to govern both effective teaching and student achievement. The selection of these "observables" seems arbitrary and atheoretical. While they covary among themselves, all too often they are but modestly related to the outcomes of interest.

Models with these properties (i.e. composed of ill-conditioned data) are often sensitive to sampling fluctuation. Their lack of orthogonality produces variance inflation of the estimates, which in turn makes them difficult to interpret.

Who would deny that teaching behaviors are responsive to the cognitive behaviors of students? If students look puzzled perhaps the teacher will give an additional example or go over the material a second time. Yet, paradoxically, models of teaching effectiveness assume that this never happens. Only additive, linear and recursive models have been estimated which assume that students are responsive to professors but not the other way round. Adequately specified teaching effectiveness models have to be nonrecursive. They have to allow for feedback parameters, but such has never been the case.

Three additional matters were addressed by the Committee members but for various reasons were included in appendices rather than in the body of the report. First, a respecified model of teaching effectiveness was formulated in order to show what kind of design is necessary if one wishes to address some of the outstanding research problems in this research tradition. Second, a compromise (but still controversial) solution to the problem of evaluating teaching was proposed. Third, a position paper was prepared as a tentative reply to the question: Why is the matter of faculty evaluation a pressing issue in the 1980's?

Recommendations

1. There should be no summative evaluation of the teaching performance of members of the Faculty.

   (In particular there should be

   (i) no student evaluation of faculty teaching performance,
   (ii) no administrator evaluation of faculty teaching performance, and
   (iii) no colleague evaluation of faculty teaching performance,

   designed explicitly for summative evaluation purposes.)

2. Guidelines for specimen instruments for the formative evaluation of teaching performance should be identified and made available to all Faculty.

   (It is recommended that the instruments include:

   (i) self-appraisal or self-evaluation forms,
   (ii) student rating forms or student instructional reports, and
   (iii) colleague observation guides.

   It is also recommended that opportunities be made available for faculty members to self-review video-taped classes.)

3. Faculty members should have the right to conduct formative evaluation of their teaching and to make the results of this evaluation available to the Dean or appropriate Department Head. However, faculty should not be required to collect such information nor to provide it to the Dean or Department Head.

4. That procedures for assisting Faculty members in self-evaluation be formalized in the Faculty.

   (Normally, this will include student and colleague feedback; but if the procedure is to be successful it must be conducted in a non-threatening manner.)

Supplemental Recommendations

S1. That Department Heads and the Dean give appropriate weight to teaching experience when making new appointments.

S2. That candidates for appointment be expected to provide both teaching references and academic references.

S3. That candidates for appointment be expected to give a Departmental seminar or lecture.

S4. That each Department be expected to identify the relative weight that will be given to the teaching criterion in decisions affecting tenure, promotion and merit salary increases.

S5. The weights assigned by Departments to the teaching criterion be negotiable between individuals and Department Heads.

CONTENTS

	Page
REPORT SUMMARY	
1. Terms of Reference	1
2. Why Evaluate Teaching Performance?	3
3. Formative Theory	5
Formative Theory 1: Selected Observations.	6
Formative Theory 2: Research Issues . . .	10
Formative Theory 3: Conclusions	12
4. Summative Theory	14
The Case For Summative Evaluation	14
The Case Against Summative Evaluation . . .	17
Summative Theory: Conclusions	31
5. Conclusions and Recommendations	35
A Model of Effective Teaching: Where Do We Stand?	35
The Terms of Reference	37
Conclusions 1: Formative Theory	40
Conclusions 2: Summative Theory	41
List of Recommendations	43
Conclusions 3: What Can Be Done?	45
A Comment on Sources	46
References	47
Addendum 1: The Dean's Memo	49
Addendum 2: Student Evaluation Subcommittee Terms of Reference	52

Page

Addendum 3: A Model of Teaching Effectiveness . . 55

Addendum 4: A Proposed Compromise Solution to the Problem of Evaluating Teacher Effectiveness 78

Addendum 5: Why Is the Matter of Faculty Evaluation a Pressing Issue in the 1980's? 82

1. Terms of Reference

The Committee on Faculty Evaluation chaired by Dr. P.J. Warren was appointed by the Dean, Faculty of Education on June 3, 1982, to study the problem of faculty evaluation. The Committee was asked to address two specific matters: first, the criteria which should govern the evaluation of a faculty member; and, second, the kinds of procedure which would be appropriate in order to facilitate such an evaluation. The Dean's memo which specified the terms of reference of the Warren Committee is appended to this report as Addendum 1.

The problem is almost self-evident. Faculty members, Department Heads and the Dean are from time to time expected to make decisions regarding the tenure, promotion and merit salary increases of individual faculty. In addition, Heads conduct annual evaluations of faculty. Impartial decisions, however, cannot be made in the absence of information about a faculty member's teaching, research, professional activity, and services to the university and university community.

In a memo of November 2, 1982 the Committee on Faculty Evaluation established a subcommittee on the student evaluation of faculty. The members of this subcommittee are the authors of this position paper. Their tasks were defined as follows:

(i) to summarize recent research findings regarding student ratings of faculty instruction;

(ii) to suggest the purposes for which student evaluation of faculty performance might be used;

(iii) to make recommendations about the use of student evaluation procedures; and if it is recommended that student evaluation of faculty be implemented,

(iv) to recommend appropriate procedures for its implementation.

Finally, the committee was asked

(v) to examine and assess student rating forms; and, if desirable,

(vi) to conduct research on their appropriate use.

The memo from the Committee on Faculty Evaluation establishing the Student Evaluation Subcommittee and specifying its terms of reference is attached as Addendum 2.

2. Why Evaluate Teaching Performance

Four reasons are often given for evaluating university teaching performance:

(i) to give individual faculty members feedback on their teaching primarily as an aid for improving their future teaching performance;

(ii) to give administrators information as a basis for making tenure, promotion and merit increment decisions;

(iii) to provide published information for student use in course selection;

(iv) to provide research data for institutional and educational research projects.

These purposes are not exhaustive. In some sources other reasons for evaluating teaching are mentioned. Three of these are noted:

(v) to protect students from incompetent teachers (cf. reason (iii) above);

(vi) to protect the teacher from arbitrary administrative decision-making (cf. reason (ii) above); and

(vii) to reward superior performance through the public identification of outstanding teaching.

Here, we consider the first two reasons. Each is identified as a separate theory: the formative theory and summative theory of the evaluation of teaching effectiveness respectively. The formative function of teaching effectiveness calls for the improvement of teaching through the identification

of a faculty member's teaching strengths and weaknesses. In this approach care is taken to avoid ranking faculty or making judgments about their teaching ability. In contrast the summative function calls for an assessment of teaching competence on the basis of which judgments may be made on matters pertaining to promotion, tenure, merit salary increments - even dismissal.

As would be expected, formative theory is less controversial; hence, more likely to be supported by university faculty than summative theory. Nevertheless, the distinction between formative and summative theories of teacher evaluation tend to be more analytical than empirical. This means that the two functions may be difficult to separate in practice. And this difficulty may account for the scepticism, even cynicism, on the part of professors toward practices designed to evaluate their teaching.

3. Formative Theory

The empirical literature on the evaluation of teaching is shot through with conflicting findings, but if the literature supports anything it is that teachers can improve their teaching skills if they are given constructive feedback on their strong and weak points. "Constructive feedback" is the key to understanding formative theory. An individual begins with some model of what a good teacher does in a classroom. He receives information or feedback about which characteristics of the model of good teaching seem to work for him and which seem not to work. At this point the individual is faced with several choices; namely, which characteristics to keep or promote, and which characteristics to reject or reduce on the basis of whether they worked or not. Further feedback promotes additional corrections. At each step the teacher becomes a better teacher. The greater the opportunity the individual has for practicing his teaching style coupled with regular feedback the more effective the teacher will become.

It will be the new teacher who is likely to make the greatest incremental improvements in teaching. This is because it is the new teacher who is most likely to be developing his teaching style. The recurrent and patterned feedback process is characteristic of the formative model. In the ideal case it will result in the adoption of a style of teaching which is the most effective for the individual given the individual's personality resources. The theory, then, gives credence to

the adage that "you can't teach an old dog new tricks." Because the feedback process is iterative and incremental, experienced teachers are much more likely than beginning teachers to have adopted a style which is successful for them.

Formative Theory 1: Selected Observations

On Teacher Characteristics. Formative theory is based on an individual's preliminary conception of a model of good teaching; hence, on notions of the characteristics of good teaching. The model is constrained, therefore, by the number of characteristics that are available to the individual; where by "characteristics" one is referring to the components of the individual's personality. These might include physique, appearance, speech, gestures, age, sex, humour and such like. But these qualities differ from individual to individual because personalities differ. Thus, what works for one teacher need not work for another. For example, some teachers are comfortable in an informal discussion-like class while others perform better in a formally structured lecture situation.

On Teacher Effectiveness. The outcome measure in the formative theory of student evaluation of teaching is improved teaching effectiveness. And just as different individuals have different characteristics available to them, so they have different conceptions of what is meant by effective teaching. In that effective teaching is the goal, then different conceptions of the goal will necessitate the use of different means for attaining it. In terms of Broudy's modes of teaching,

for example, one would assume that the teaching process underlying the didactic mode would be quite different from that underlying the heuristic mode. Teachers, then, have to match their personality characteristics to their notions of effective teaching. In turn, this means that they have to weigh and evaluate their personality attributes in the course of adopting an optimum teaching style, or in the course of maximizing their teaching effectiveness. The process is congruent with the assimilation and accommodation processes defined by Piaget accounting for schema formation or frame construction. In this instance the frame we are referring to is the individual's professional conduct and teaching style.

In general the teacher's selection of a style or mode of teaching will stimulate the formative process for achieving the style. The more effective the teacher's use of his personality resources for achieving a desired style the greater the probability of obtaining positive student evaluations. Even so, there can be no guarantee that a positive evaluation will be synonymous with teacher effectiveness where by teaching effectiveness one is referring to maximization of student learning. The definition of effective teaching which results in the maximization of the student evaluation of teaching performance is not one that will necessarily guarantee the maximization of student learning (Lima, 1981: 1056).

On Teacher Utility. The maximization of teaching effectiveness via a formative model of teacher effectiveness is still only a subsidiary component of the overall problem; namely, that of maximizing professor utility. The utility function only depends in part on teaching effectiveness. In addition factors such as research, administration and committee work have to be taken into consideration.

Recently, John Centra, who is arguably the foremost authority on the evaluation of faculty performance has listed those factors that are most commonly used in evaluating overall faculty performance (the utility function). The seven most important factors in order of importance are classroom teaching, quality of publications, number of publications, personal qualifications (honours, awards, degrees, etc.), research, supervision of graduate student research and committee work. In relation to these factors activities in professional societies, student advising, community service, consulting and competing job offers were considered to be either minor factors or of negligible importance.

But these utility functions will, like teacher characteristics and perceptions of teacher effectiveness, differ from individual to individual. The individual who places emphasis on the teaching factor is likely to adopt a different style of teaching than the professor who places emphasis on research. In view of this possibility it becomes essential that faculty members be aware of the weighting schemes used by

tenure and promotion committees because these may have a bearing on the type of style adopted. It is not unreasonable to expect an individual's teaching style to be influenced by their expected institutional rewards.

What do these observations mean in terms of formative theory? They all support a model of teaching effectiveness in which different teachers are likely to have different modes of teaching. The formative theory does not represent so much a model of teaching as a family of models. The characteristics selected in order to optimize effective teaching will, therefore, differ from one individual to the next depending on the style preferred. The style adopted may also depend on the teacher's conceptions of teaching effectiveness. Teaching effectiveness, in turn, may be influenced by the weighting schemes used by tenure committees. It is conceivable, for example, that two individuals with roughly equal personality resources would develop different modes of teaching based on the expected rewards at their respective universities. A strong research orientation would likely promote a heuristic teaching style while a strong liberal arts orientation might favour a human relations style. Thus, one might expect aggregate teaching styles in Ph.D. granting research universities to be different than those in four year liberal arts colleges.

Formative Theory 2: Research Issues

The formative model of teaching effectiveness is fraught with technical difficulties. Three will be mentioned; namely, the problem of measuring change, the problem of nonlinearity and the problem of measuring utility function constraints.

Measuring Change. In terms of formative theory an individual's teaching effectiveness changes over time as a function of constructive feedback. The model, then, is dynamic, rather than static. Yet, of the more than two thousand references that Murray has located in the literature on teaching evaluation over the past 15 years only a handful - less than one percent - have been studies of the dynamics of teaching effectiveness. Researchers in education have found it more convenient to hypothetically stop the moving system in order to examine its cross-sectional features rather than to look at time-series data which would permit the researcher to examine changes in the magnitude of variables over time. Effectively, this means that formative models of faculty teaching effectiveness have not been formulated, and neither have they been estimated. We have developed our models of teaching by looking at static models when we should have been looking at dynamic models instead.

Nonlinearity. Consider a personality characteristic such as "sense of humour" which in theory should contribute positively to effective teaching. The fact is that humour,

like the exclusive use of positive sanctions (rewards), will probably only contribute positively up to a certain level. Beyond a certain point (of inflection), however, humour may detract from effective teaching. While jokes up to a point will do no harm, the teacher who spent most of his class time telling jokes is unlikely to enhance his teaching effectiveness. The teacher who places excessive dependence on only positive reinforcement or who sets unrealistically high expectations or who overuses a normally positive characteristic such as empathy may find his strategies backfiring. These are teaching strategies which are likely to have non-linear relationships to the preferred mode of teaching. Yet, no formative model of teaching effectiveness has yet incorporated the polynomial terms necessary to test a nonlinear model.

Utility Function Constraints. It was noted above that utility functions will differ from individual to individual. If one professor emphasizes teaching and another publication, they are likely to adopt different teaching styles. Yet seldom is the teaching orientation or any other single orientation an exclusive criterion for career progress decision making. Recently Rogers noted that "typically, in making these decisions greater and equal weight is given to teaching and research, and less to service (e.g., 2:2:1 for teaching research and service respectively)." Time spent on teaching will be constrained by time spent on other factors in the utility function equation. The factors, however, will not be completely independent of one

another even though linear independence is a fundamental assumption underlying the accurate estimation of linear equations. The degree of departure from orthogonality or linear independence will determine the extent to which the estimates will be inflated and it is difficult to interpret inflated estimates with any precision. There are strategies for coping with this problem of nonorthogonality such as the use of principal components or factor analytic aggregation of variables and ridge regression but in the teacher evaluation research tradition these methods seem to remain unknown, therefore, unused.

Formation Theory 3: Conclusions

Three conclusions can be drawn from this analysis of the formative model of teaching effectiveness. In the first place we believe that most teachers can improve their effectiveness if they are given constructive feedback and the opportunity to develop a style which is congruent with their personality characteristics. In other words we believe that the model is a viable one which goes some of the distance toward accounting for good teaching. In the second place the theory does not so much promote a model of teaching effectiveness as a family of models. This is because there is no reason to believe that different teachers will have the same teaching style. It is because different styles are effective for different individuals that correlational studies of the student evaluation of teaching

performance generate only small coefficients, and why therefore the results are so insignificant and inconclusive. In the third place, because the formative model is a dynamic rather than static one; hence, cannot be captured by the traditional cross-sectional design, it is one fraught with technical difficulties. Though these can be overcome there have to date been no time series studies of the determinants of teaching effectiveness. Thus, the formative model to data remains effectively unexamined and untested.

4. Summative Theory

Summative theory is supportive of the systematic, formal evaluation of teaching performance. It is based on the assumption that the scientific method is an effective procedure for measuring human behavior. As will be discussed below, this assumption seems premature. Teaching behaviours at the present time cannot be measured with much precision.

The Case For Summative Evaluation

1. Research into the processes of teaching evaluation is timely. Policy statements in negotiated agreements between teachers and school boards and between university faculty and boards of governors are to be found which make some form of evaluation mandatory. The revised draft of the terms and conditions of employment for faculty at Memorial University, for example, directs a Department Head to "conduct an annual evaluation of each faculty member under his/her administrative jurisdiction...." These terms and conditions also identify five criteria which are "normally to be taken into consideration." One of these is teaching. There seems little doubt that faculty evaluation is here to stay.

2. Universities through their charters and constitutions are enjoined to give instruction in every branch of knowledge and learning. Pursuant to this mandate university authorities take their teaching function seriously; usually by holding their senior appointees accountable for the quality of teaching.

3. The Constitution of Memorial University defines its functions as follows: "... to establish and maintain... faculties...departments...to give instructions and training; to grant degrees, ...to conduct...research...." In other words the charter places teaching on an equal footing with research. By rewarding faculty members on the basis of their teaching as well as on the basis of their research and scholarly publication the University meets the spirit of the charter. The public relations value of promoting excellence in teaching should not be overlooked, even though formal procedures for giving recognition to teaching excellence at Memorial are not as well developed as elsewhere.

4. A fourth argument holds that the use of teaching evaluation data for summative purposes provides the incentive for faculty members to take the time and effort necessary to excel in teaching. It may be unrealistic to expect excellence in teaching in the absence of powerful incentives such as

salary increase, promotion and tenure.

5. A fifth argument is predominantly economic. As tuition costs escalate, students and their parents are increasingly demanding proficient instruction as a minimal return on their investment. Concomitantly, cut-backs in university operating revenues have pressured university administrations into the adoption of selective tenuring and tight promotion policies. Thus, competition for increasingly scarce rewards intensifies - a situation making personnel decision-making more difficult. One consequence of economic constraints has been greater demand for explanations regarding personnel decision making; which, in turn, calls for more information on faculty teaching performance.

6. Murray (1980) posits a sixth advantage for the summative evaluation of faculty teaching; namely, its compensatory function. Personnel decisions affecting promotion, tenure and salary increments are often governed by performance in the single area of research and publication. The absence of objective information on teaching results in this criterion being ignored. The summative evaluation of teaching, however, makes it possible for a faculty member to gain tenure and promotion through performance in two areas - teaching as well as research. Thus, argues Murray, a majority of faculty prefer an evaluation system where modest performance in one area may be compensated for by good performance in another. Few faculty wish to put all their eggs in one basket.

7. Kronk and Shipka (1971) in their National Education Association handbook for faculty leaders discuss the evaluation of faculty teaching. They note that faculty members have always competed for scarce rewards, hence, the desirability of recurrent faculty evaluation. But when faculty are not evaluated on a scheduled basis the door is left open "to administrative decision making which need not reflect systematically gathered data on faculty performance." Systematic evaluation restricts "managerial prerogatives", and may deter an administrator from a course of action which could not be justified in the event of an appeal. On these grounds the NEA handbook concludes that "the best approach is to require evaluation of all faculty on a scheduled basis."

The Case Against Summative Evaluation

The Means/Ends Correspondence Dilemma. It would seem incumbent upon the advocates of the formal evaluation of teaching to show that there is a significant degree of association between teaching evaluations and the improvement of teaching. There is no such evidence. Similarly, there is an absence of unambiguous evidence that positive evaluations of teaching lead to significant positive changes in student learning.

The lack of correspondence between teaching evaluations and student learning is odd. Two observations on the dilemma may be worthwhile. In the first place efforts to measure teaching effectiveness on the basis of its impact on

student learning are unpopular. This is because student learning is a multicausal activity. Many of the significant factors affecting student learning - family background, motivation, aptitude, study habits, course load, class size - are outside the control of the teacher. Thus, to hold teachers responsible for the learning behaviors of their students is unreasonable. University students are treated as autonomous persons. Just as parents are not held responsible in the de jure sense for the criminal behaviour of their children, so teachers should not be held accountable for the out of class behaviours of their students.

In the second place, there seems to be tacit understanding of this dilemma on the part of administrators and colleagues who evaluate faculty performance. In a recent study (Gosling 1978) over 80 percent of the 2,202 comments that administrators made on rating forms were found to be positive, whereas only 6 percent were of the "needs improvement" variety. Gosling accounts for this on the grounds that administrators find reasons outside the classroom context to account for observed weaknesses. In other words the observed limitations are accounted for in terms of factors which are external to the institutional setting.

Theoretical Naivete 1: Dimensions of Teaching. Theories of effective teaching are underdeveloped. For example, few studies provide a theoretical rationale for selecting the dimensions of teaching which are included in

teacher evaluation rating scales. A host of political, pragmatic and parochial factors are involved in deciding which dimensions of teaching to include and which to exclude.

When the writer began school teaching in the mid 1950s a primary concern of teacher educators of the day was teaching for democracy and the democratic way of life. Today, this theme if not abandoned is dormant. When the writer began university teaching in the 1960s a demand of the student movement of the day was student participation in university government. Faculty in the arts and social sciences in particular would consult with their students in advance about course grades, examinations and even course content. Today few rating forms refer to the principle of student participation. There seems little theoretical basis to many teacher evaluation instruments, which are more likely to reflect ephemeral matters rather than seeking for more enduring, invariant relations.

Theories of teaching do exist. For example, on the basis of social learning theory one could deduce important dimensions of the classroom environment: the stimulus or opportunity to learn dimension, the reinforcement or reward/punishment dimension, the expectation/encouragement dimension and the role model dimension. A structural-functional characterization of teaching could be based on the four functional imperatives inherent in any action system such as a school classroom: adaptation, goal attainment, integration

and pattern maintenance or latency. These have been translated into the domains of the classroom environment by Australian researchers Williams and Batten (1980); namely, the opportunity domain, the adventure domain, the identity domain and the status domain respectively. The teacher who meets these personal fulfillment expectations of students is likely to be an effective teacher.

It is because theories such as these have not been used to facilitate the formulation of models of teaching that current research efforts appear to be atheoretical. And whether these theories are competing or complementary entities remains unknown.

Theoretical Naivete 2: The Problem with the Holistic Approach. There are numerous studies which have examined the relationship between teacher ratings and student achievement. Correlations in these studies vary widely - between +.75 and -.80. Abrami et al. (1982:458) account for these inconsistent findings on the grounds that different investigators use different research designs. Though the designs of some studies are flawed several correlational studies (e.g., Centra, 1977; Ellis and Rickard, 1977; Shavelson, et al., 1976; and Sullivan and Skanes, 1974) have been identified as having "strong" designs. In each of these strong design studies, however, the assessment of teaching performance was based on a single question on the rating scale. This item calls for the evaluator to make an overall assessment of teaching ability in terms of an order of "goodness" from most desirable

to least desirable. The usual labels include: outstanding, excellent, above average, below average, needs improvement, poor, or some variant thereof. This type of global item, however, does not represent a dimension of teaching. The snag is that holistic assessments because they are based on a single item are likely to be biased. And since the reliability coefficients of such variables cannot be estimated it is not possible to estimate their bias or to correct for bias.

Researchers using holistic measures of teaching effectiveness are usually able to show that they correlate more highly with student learning than do specific ratings of teaching behaviour. And this may account for the reason why holistic evaluations are the ones most frequently used for making critical personnel decisions about promotion and tenure. These investigators are well aware of the psychometric limitations of their instruments; thus, researchers such as Sullivan (in a recent Faculty seminar) recommended that holistic variables be used solely in formative evaluations.

Formative evaluation calls for the improvement of teaching through feedback processes whereby a teacher's strengths and weaknesses are identified. Given this purpose it is difficult to say what the relevance is of overall assessments of teaching ability. An holistic assessment does not represent a component or dimension of teaching. Because such information does not relate to the particular strengths or weaknesses of teaching, from the point of view of formative evaluation it is redundant information.

Theoretical Naivete 3: Model Specification. In this section of the discussion of the case against summative evaluation we draw attention to the misspecification problems of the models of teaching effectiveness. Models may be misspecified because important structural components should have incorporated but which were erroneously excluded, or because components were erroneously included but which should not have been, or because the model was functionally or mathematically misspecified.

The first specification problem to be addressed will be referred to as the problem of generalizability. The dimensions of teaching which are often found on teacher evaluation rating instruments usually include the following: organization, discipline, atmosphere for learning, sense of humour, health and vitality, appearance, enthusiasm, handling routine procedures, and communication skill. Appropriate distractors such as strongly agree through strongly disagree or outstanding through needs improvement are attached to each dimension. It is assumed, then, that each dimension is an observable phenomenon. Theory at this level of analysis is concerned largely with the explanation of relationships between sensory data or observables.

Analysts wish to know two things: first, the magnitude of the relationship between each dimension of teaching and a faculty member's overall teaching performance or teaching style; and, second the magnitude of the relationships between these observables and student achievement.

According to the theory a well organized teacher, in control of a well disciplined class, with a sense of humour and with the ability to communicate his enthusiasm for the topic is likely to be given a high rating by students. And by the same token the highly rated teacher is likely to stimulate high student achievement other things equal.

But what about lecture planning, clarity of lecture presentation, rapport with students, use of instructional aids (the blackboard, models, video-tapes), punctuality, and instructional skill? The number of possible dimensions one could add to the list of dimensions could go on and on. And in the absence of an articulated theory of teaching there would seem to be no framework for governing the list of potential dimensions.

Explanation in the form of observables has several advantages. First, such explanations are not far removed from conventional wisdom explanations. They are, therefore, intuitively appealing. Second, empirical explanation is usually easily tested; and especially at the bivariate level. Third, empirical explanation is practical in the sense that the practical implications of the findings usually flow directly from the findings themselves. Three conclusions are often drawn from the findings of teacher evaluation studies where explanation is in terms of observables. First, the correlations between the dimensions of teaching and the student outcomes in terms of achievement are usually modest. Second, the correlations between the dimensions of teaching

themselves are often substantial - well organized teachers tend to be enthusiastic teachers and so on. Thirdly, while teachers who have received a low summative rating on a particular dimension can usually increase their ratings on the dimension in question, it is seldom that these efforts have any significant effect on their overall global ratings. The teacher, for example, who is rated as mediocre on "examination question format" is usually able to write better tests the next time round. But the improvement is unlikely to affect a significant change in an overall rating.

It seems, then, that while empirical explanation is easily tested the findings are sometimes not very much help. One solution to the problem calls for the rejection of empirical explanation in terms of observables, in favour of conceptual explanation in terms of unobservables, latent constructs or theoretical concepts. Empirical explanation is usually specific to a particular situation whereas conceptual explanation is intended to apply to all situations which fall within the scope of the explanation.

In recent years models formulated at the level of empirical explanation have been criticized for their concreteness; that is, for the fact that they are too situation specific. The argument invoked is that the principle of generalizability in social research overrides the principle of inclusiveness or (misplaced) concreteness. In other words, it is pointed out that the purpose of modeling is not to mirror social reality, but rather to reduce that reality to a set of fundamental and more manageable features.

Models concerned with the specification of relationships between theoretical (nonobservable) concepts are referred to as structural models. Statistical models isomorphic to structural models have been formulated by Joreskog and Sorbom (1978). Kerlinger (1982) notes that estimation of these structural equation models is dependent on the analysis of their covariance structure - a mode of analysis which is a far cry from correlational and multiple regression models of the naive empiricists. It is to be regretted, therefore, that the structural equation approach to the estimation of models of teaching effectiveness has never been tried.

A second misspecification error in models of teaching effectiveness will be referred to as the Jackson effect. At a recent Faculty seminar given by Dr. Arthur Sullivan, Michael Jackson commented on the fact that he often found that the responsiveness of students to his instruction differed from class to class. He noted that this was the case even in different sections of the same course and even though the university registration procedure was supposed to ensure the random distribution of students between course sections. If the standard explanations of teaching effectiveness are correct one should find that Jackson's teaching would have the same effect on one class as the next, when the classes are in the same course.

What the traditional model fails to acknowledge, however, is that students are not peices of clay which

can be molded at will by an instructor. Rather, students have the capacity to influence the instructional environment presumably to their own advantage. And this capacity will vary somewhat from class to class. If Jackson has a faculty member auditor in a class he may respond differently to that class than to a class which does not have a faculty member auditor, but which may have an older student who is on educational leave from his or her Board.

Because students have the capacity to influence instruction by prompting the instructor, by asking or not asking questions, by participating or not participating in discussion, the flow of influence in the classroom is not one-way from instructor to student, but two-way from instructor to student and reciprocally from student to instructor simultaneously. Technically, the one-way flow of effect model is known as the recursive model. But this model, because it would not incorporate the Jackson effect would be seriously misspecified. The correct model specification should be a two-way, nonrecursive or reciprocal model. The Jackson effect calls for the estimation of reciprocal or feedback relationships between the class environment and student cognition, and for the estimation of the responsiveness of student achievements to both the structure of the class environment and the cognitive abilities of students. It is to be regretted that feedback models of this kind have never been estimated by teacher evaluation researchers.

A third specification error in models of teacher evaluation concerns the fact that where students are used to rate teaching effectiveness they do so usually toward the end of a course of instruction. By this time students have received feedback from lab reports, mid-term tests and the like about their course performance. Does this prior knowledge bias student ratings? Will students who are performing well in a course give a weak professor a low rating and vice versa? If not, then the observed relationship may have a tautological component to it.

In order to take the tautology hypothesis into account the outcome variable would have to be a time two minus time one achievement change score, and time one achievement would be used as a covariate or control variable. Such a design is longitudinal, involving data gathering at at least two points in time. The writer knows of no such study in the teacher effectiveness research literature.

The Dr. Fox Effect. Several observers of the teacher evaluation research literature have noted that the very visit of an evaluator to a classroom will alter the teaching. Thus, what an evaluator observes need not represent typical day to day teaching. To demonstrate this susceptibility of observers to variations in teaching behaviors lecturers were trained to be slick, enthusiastic and dynamic, while at the same time teaching nonsense. In several replications of the "experiment" these "teachers" were able to generate high marks from students and peers who were asked to rate their performance. This bias

in the ratings is known as "the Dr. Fox effect". It may be found in almost any situation in which a teacher is being assessed.

Observer Bias. Scriven (1980) - former President of the American Educational Research Association - argues that the summative evaluation of teaching is not only wrong but "a disgrace." He points out that every observer brings to the evaluation situation a set of implicit and explicit assumptions about teaching that will influence the interpretation he places on what is observed. In short, the observer influences the process.

Recently, Art Ponder of the Department of Educational Administration has conducted some ingenious research designed to identify observer bias. First, he identified the dominant teaching orientation of a large number of school administrators in terms of Broudy's triad - philetics, didactics and heuristics. Each administrator was next asked to rate three hypothetical teachers in terms of their teaching effectiveness. As the reader will have inferred, one teacher was described in terms of a philetic orientation, one in terms of a didactic orientation and the third in terms of a heuristic orientation. Ponder's hypothesis was that if there were observer bias, philetic administrators would rate philetic teachers more highly than didactic or heuristic teachers; and, similarly, didactic administrators would rate didactic teachers highest, and heuristic administrators would favour heuristic teachers.

His hypothesis was strongly confirmed. An interesting finding was that the administrators were almost evenly divided between the three orientations.

Bureaucratic Effects. In this writer's judgment the consequences of accepting a summative model of teacher evaluation may be the most important reason for deciding against it. University administrators are the ones who will be using summative evaluation of teaching information for decision-making purposes. They are cognizant of the sensitivity of university faculty toward judicial matters and of the litigious spirit permeating faculty associations during a period of economic retrenchment in university affairs. This spirit is congruent with the times - times in which the recently patriated Constitution with its entrenched Bill of Rights is being invoked to solve a host of outstanding legal issues from native rights to women's rights.

It is not sufficient for an administrator to discern that a faculty member needs to improve, or that his teaching is weak. It is necessary to be able to show objectively that such a judgment is impartial, that the grounds for the judgment are valid. The National Education Association in the United States states that due process requires the provision of reasonable safeguards to protect the legitimate rights and interests of faculty members. Their list of safeguards include the following.

(i) An individual should receive specific, valid reasons for an adverse decision.

(ii) An individual should receive prompt notice if his teaching performance is evaluated as deficient in some respect.

(iii) Prior to an adverse decision the individual should be given a reasonable length of time to correct the noted deficiency.

(iv) Institutions using summative methods of evaluation should provide a program designed to assist faculty members who wish to improve their teaching performance.

(v) Faculty member participation in such a program should be voluntary.

(vi) All information on a faculty member's teaching performance should be accessible to him.

(vii) Faculty members should have the opportunity to respond in writing to an evaluation of his teaching performance; and, in particular, should be able to request information on the accuracy (reliability), relevance, completeness, validity and consistency of the evaluation.

(viii) An individual should have an opportunity of purging any institutional file of prior teaching evaluations after a specified time and especially when allegations in any evaluation are inaccurate, irrelevant or incomplete.

(ix) Any evaluation which results in an adverse personnel decision such as nonrenewal of contract, tenure denial, promotion denial or the withholding of merit salary increments, should have the right to appeal the decision to an impartial source.

(x) In such an appeal the individual should have the right to representation.

(xi) The standards of performance of a faculty member should be made explicit at the time of appointment and should remain constant unless the individual and an appropriate administrator agree to change the standards.

(xii) Faculty members should be evaluated impartially; that is, should be evaluated in the same fashion as other faculty members who do the same kind of work. This means that members of the Department of Curriculum and Instruction, for example, need not be evaluated using the same criteria and standards as members of the Department of Educational Psychology.

But such safeguards, if they are considered necessary, are costly, and demanding of an administrator's time and energy. The system would need to be monitored. This would involve adding another bureaucratic peg to an already over-bureaucratized institution. The NEA concludes (Kronk & Shipka, 1981) that few arrangements for monitoring the safeguards for ensuring due process "will prove ... as effective as a faculty union which negotiates and enforces master contracts skillfully."

Summative Theory: Conclusions

The summative function of teaching effectiveness calls for the formal assessment of a teacher's teaching ability. As was explained above summative evaluation is here to stay. The evaluation of University faculty is called for in the terms and conditions of the employment of faculty. University constitutions enjoin faculty members to take their teaching functions as seriously as their research functions; therefore, most universities provide incentives for faculty members to take the time and effort necessary to become good teachers. Summative evaluation of teaching is also advantageous for the teacher in that it provides protection against arbitrary personnel decision making.

In theory, then, it is not difficult to support a summative function of teacher effectiveness. In practice, however, there are numerous problems with the implementation of summative methods of evaluating teaching. The evidence

that good teaching necessarily promotes high achievement is lacking. Much of the variance in achievement is attributable to forces beyond the control of the teacher. Administrators are reluctant to evaluate a faculty member's teaching performance and when they do have a tendency to inflate their assessments. The student evaluation of teaching, then, is a "cop out" solution. One reason for the reluctance of administrators to evaluate teaching is that there is an absence of a clearly articulated theory of teaching. (But see the section on formative theory.) Thus, the selection of the dimensions of teaching for inclusion on rating scales is arbitrary and subject to the fads and fashions of the day.

Much of the research on teaching has focused on observables or easily measured variables such as overall assessments of teaching ability which are subject to bias and measurement error. The research literature describes numerous models of teaching which are seriously misspecified. For one thing the models contain too many dimensions of teaching which are highly correlated with one another, but related only modestly to teaching effectiveness. This means that the models lack orthogonality, are subject to sampling fluctuation and, in general, account for only modest variance in teacher effectiveness and student achievement. For another, the models fail to account for the reciprocal or feedback relations between teacher behavior and student behavior in the classroom. In other words, how a teacher teaches will depend on the responsiveness of the student, and the responsiveness of

the students will depend in turn on the teacher's behaviour. This interdependency - though obvious to the most casual observer - calls for the estimation of feedback models. But a feedback model of teaching has never been formulated or estimated in the enormous literature on the evaluation of teaching.

Another specification error which is found in some of the most frequently cited literature concerns the failure to control for the fact that student assessments of their professors are usually requested near the end of a course of instruction, by which time students often know how they are being evaluated by their professors. It is believed that this prior knowledge of how well a student is performing in a course could bias his judgment; hence, if not controlled for, could account for a tautological element in the model. And if this were not enough there is considerable evidence that observer bias may seriously undermine the impartiality of teacher assessments.

The most problematic feature of summative evaluation lies, however, with its bureaucratic consequences. Universities committed to the use of summative methods of teacher evaluation whether by student ratings or administrator ratings would have to be able to guarantee due process through the provision of safeguards. This would promote the formalization of the process, or the establishment of self-rigidifying structures which are likely to prove resistant to change. Such processes have a tendency to counteract their own purposes - to stultify

good teaching instead of promoting it. Teachers subject to the formalities of student or administrator assessment may play it safe. Some will be inhibited by the procedures, others will strive, like Dr. Fox, for effect. Few will feel free from system constrains even though such freedom may be an essential prerequisite to outstanding teaching.

5. Conclusions and Recommendations

A Model of Effective Teaching: Where Do We Stand?

We do not have a well-specified model of effective teaching despite the quantity and high quality of research effort in the past decade. The most important problem facing the researcher is the fact that different teachers do not necessarily adopt the same teaching mode. As a result there is a mismatch between teacher characteristics and competencies on the one hand and the usual measures of teaching performance on the other, which accounts for the insignificant correlations between predictor variables and performance outcomes in much of the research literature.

The problem is widely recognized. Thus, a host of studies have been mounted designed to identify effective teaching styles (Kleine, 1982). Many of the studies have been pursued within the well-known process-product framework. Several ambitious studies have been conducted within this framework but with the additional complication of considering interaction effects. These latter studies have tried to identify the most appropriate instruction for different types of learner; hence, the search for aptitude treatment interaction (ATI).

Most of the teaching style studies have been concerned with identifying the style, or a style for all seasons. In general, the research has failed in its objective. Early studies focused on dominative vs. integrative, teacher

centered vs. learner centered, and authoritarian vs. democratic styles. In the past decade research on direct vs. indirect, traditional vs. progressive, formal vs. informal, and open vs. closed styles has been pursued. Though the assumption has been that these styles are different from one another one wonders whether "a rose by any other name would smell as sweet." In any event the style for all seasons is as elusive as ever.

Where do we stand? The search for the relationships between teacher styles and student achievement is not in vain, what researchers are beginning to conclude is that the effects of teacher style are probably not direct because they are mediated by two sets of intervening variables; namely the classroom environment on the one hand and the comprehension or cognitive strategies of the learner on the other. Thus, pupil achievement is hypothesized to be highly responsive to both intervening factors; namely the influences stemming from the learning environment itself and the learning strategies adopted by the students. It can be assumed on the basis of the earlier discussion on feedback effects that these two sets of intervening variables will be reciprocally related; that each affects the other.

These intervening constructs, in turn, are probably highly responsive to both the performance styles of the teacher and the teacher personality characteristics or teaching resources. A schematic model of these hypothesized relationships is depicted in Figure 1. What we are assuming,

therefore, is that while the direct effects of teacher styles on student achievement may be modest (e.g., the zero-order relationships) the indirect effect and hence their total effects may be substantial. This is because teacher styles have powerful direct influences on students' comprehension strategies and the classroom climate, which in turn directly affect student competencies.

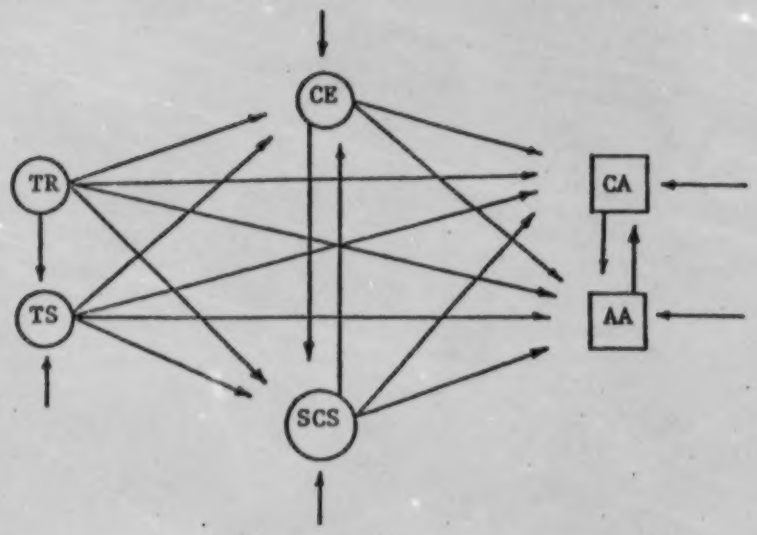
Figure 1 About Here

The complexities of the model depicted in Figure 1 need not concern us here. The interested reader is referred to Addendum 3 for the formulation of a structural equation model of teaching effectiveness.

The Terms of Reference

Caveats. Though this report was commissioned as a study of "Student Evaluation of Faculty Teaching", there is little reference to "student evaluation" in the discussion. Similarly, there is little reference to "Faculty". This is because the arguments presented may refer just as validly to administrator evaluation or colleague evaluation as to student evaluation. By the same token the observations reported above refer as legitimately to school teachers as they do to university faculty. We only resort to the use of the delimiter "student evaluation" when the discussion is relevant for this level of analysis.

Figure 1: Schematic Diagram of a Teaching Effectiveness Model



Key: TR = teacher resources, TS = teaching style, CE = class environment, SCS = student's cognitive style, CA = cognitive achievement, AA = affective achievement.

Recent Research Findings. The literature is so vast that the task of reviewing the recent research findings within the time frame was not feasible. Instead, we relied on several recent reviews of the related literature. These are identified below in the section on "Sources and Additional Reading".

Purposes of Student Evaluation. Four major reasons were identified for evaluating university teaching performance. See section 2 of the report. Formative evaluation is based on constructive feedback designed to help the individual teacher to develop an effective teaching style. Summative evaluation provides administrators with information which can be used in personnel decision making; for example, for making tenure, promotion and merit increment decisions. These two purposes were described in some detail in sections 3 and 4 of the report. Two other reasons were identified but not discussed, namely, to give students information on which to have course selection, and to provide research data in educational research projects.

Recommendations. The Committee formulated four major recommendations and five supplemental recommendations. Since summative evaluation is controversial and inadequately conceptualized it was not recommended as a procedure for evaluating faculty teaching performance. In contrast, formative evaluation was found to be a valuable procedure for the improvement of teaching; especially early in one's teaching career. It was recommended, therefore, that the Faculty develop guidelines and identify specimen instruments that could be brought to the attention of interested faculty members.

Outstanding Terms of Reference. The Committee was also asked to recommend procedures for the evaluation of teaching; and to examine and assess student rating forms. Both terms of reference were discussed. But the Committee members regarded them as redundant in view of the decision not to recommend implementation of summative methods of evaluation. See the list of recommendations below.

A final term of reference - to conduct research on the appropriate use of student rating forms - was also discussed by the Committee. Given the several limitations of the summative evaluation model of teaching effectiveness the design of a reliable and valid student rating form would be a valuable research project. As discussed in Addendum 3 such a project might best be conceived in terms of a larger study designed to test a structural equation model of teaching effectiveness.

Conclusions 1: Formative Theory

Three conclusions were drawn from the analysis of the formative model of teaching effectiveness. First, and most important, teachers can significantly improve their effectiveness if they are given constructive feedback and the opportunity to develop a teaching style which is congruent with their personality characteristics. Second, formative theory promotes a family of models because different teachers adopt different teaching styles. Third, the formative model

is a dynamic rather than a static one. Because of this there are several technical difficulties which have to be coped with if it is to be tested. These difficulties can be overcome but in effect the formative model still remains an untested theory.

Conclusions 2: Summative Theory

There can be no doubt that summative evaluation of teaching performance is here to stay. University charters enjoin faculty members to take their teaching functions as seriously as their research functions. Thus, universities provide incentives for faculty members to take the time and effort necessary to become good teachers. Summative evaluation is advantageous for faculty because it provides protection against arbitrary personnel decision-making.

In theory, then, it is not difficult to support a summative function of teacher effectiveness. But in practice the implementation of summative evaluation is fraught with difficulties. For example, there is little firm evidence that good teaching necessarily promotes student achievement. This is partly because much of the variance in teaching is attributable to factors other than teaching and partly because few well designed studies of the relationship between teaching and learning have been conducted.

For example, there does not seem to be a clearly articulated theory of teaching as indicated by the absence of a rationale for the selection of the dimensions of teaching included on extant rating scales. What models have been

formulated and estimated are often seriously misspecified. They focus on observables, hence, contain far too many dimensions of teaching. These dimensions though highly intercorrelated are but modestly related to teaching effectiveness. Thus, the models lack orthogonality, are subject to sampling fluctuation and account for only modest variance in either teacher effectiveness or student achievement.

The models also fail to incorporate the necessary feedback relations between teacher behaviors and student classroom behaviors. Prior knowledge of how well a student is performing in a course is seldom controlled for and neither is observer bias.

The most problematic feature of summative evaluation, however, lies with its bureaucratic consequences. In order to justify the use of summative methods of teacher evaluation administrators have to be able to guarantee due process. But the very existence of bureaucratic arrangements designed to guarantee due process tend to counteract their own purposes. The machinery can inhibit good teaching because it constitutes a set of formal constraints on the freedom to teach.

RECOMMENDATIONS

1. There should be no summative evaluation of the teaching performance of members of the Faculty.

   (In particular there should be

   (i) no student evaluation of faculty teaching performance,
   (ii) no administrator evaluation of faculty teaching performance, and
   (iii) no colleague evaluation of faculty teaching performance,

   designed explicitly for summative evaluation purposes.)

2. Guidelines for specimen instruments for the formative evaluation of teaching performance should be identified and made available to all Faculty.

   (It is recommended that the instruments include:

   (i) self-appraisal or self-evaluation forms,
   (ii) student rating forms or student instructional reports, and
   (iii) colleague observation guides.

   It is also recommended that opportunities be made available for faculty members to self-review video-taped classes.)

3. Faculty members should have the right to conduct formative evaluation of their teaching and to make the results of this evaluation available to the Dean or appropriate Department Head. However, faculty should not be required to collect such information nor to provide it to the Dean or Department Head.

4. That procedures for assisting Faculty members in self-evaluation be formalized in the Faculty.

   (Normally, this will include student and colleague feedback; but if the procedure is to be successful it must be conducted in a non-threatening manner.)

Supplemental Recommendations

S1. That Department Heads and the Dean give appropriate weight to teaching experience when making new appointments.

S2. That candidates for appointment be expected to provide both teaching references and academic references.

S3. That candidates for appointment be expected to give a Departmental seminar or lecture.

S4. That each Department be expected to identify the relative weight that will be given to the teaching criterion in decisions affecting tenure, promotion and merit salary increases.

S5. The weights assigned by Departments to the teaching criterion be negotiable between individuals and Department Heads.

Conclusions 3: What Can Be Done?

On the one hand the evaluation of faculty teaching is called for in terms and conditions of employment, on the other hand this committee cannot recommend summative methods for evaluating teaching effectiveness for the reasons given. What can be done? A possible alternative to formal summative methods such as student evaluation, administrator evaluation or colleague evaluation was discussed by the committee. It could be referred to as a compromise position or a "soft-nosed" option. Even so, the committee members were unable to agree on a compromise solution. For this reason it is presented as Addendum 4.

A Comment on Sources and Some Suggestions for Further Reading

Sources. The discussion on "Formative Theory" benefitted from an analysis of teaching effectiveness by Lima (1981) which appeared in the American Economic Review. Several ideas developed in the section on "Summative Theory" were found in articles by Kronk and Shipka (1981) and Raths (1982). Three articles in the Encyclopedia of Educational Research, Fifth Edition (1982) are recommended as useful background reading; namely, "Evaluation of Teaching", "Teacher Effectiveness", and "Teaching Styles".

Further Reading. There have been over 2000 studies on the evaluation of teaching alone in the past ten years. Fortunately it is a field which is well covered by review articles; most of which are of recent origin. The following review articles are highly recommended: Abrami et al., 1982; Klein, 1982; Knapper, 1981; Medley, 1982; Murray, 1980; and Raths, 1982.

REFERENCES

Abrami, P.C., Leventhal, L., and Perry, R.P. Educational seduction. Review of Educational Research, 1982, 52, 446-464.

Cartter, A.M. The supply and demand of college teachers. American Statistical Association Social Statistics Proceedings. Washington, D.C.: American Statistical Association, 1965: 70-80.

Centra, J.A. Student ratings and their relationship to student learning. American Educational Research Journal, 1977, 14, 17-24.

Centra, J.A. Determinants of faculty effectiveness. San Francisco: Jossey-Bass, 1980.

Ellis, N.R. and Rickard, H.C. Evaluating the teaching of Introductory Psychology. Teaching of Psychology, 1977, 4, 128-132.

Gosling, A.W. Teacher evaluation: An examination of formal expectations and actual content (Doctoral dissertation, Indianna University, 1978). Dissertation Abstracts International, 1979, 39, 5228A.

Joreskog, K.G. and Sorbom, D. LISREL IV. Analysis of linear structured relationships by the method of maximum likelihood. Chicago: National Educational Resources, 1978.

Kerlinger, F.N. The future of educational research: Part II. Educational Researcher, 1982, 11:9, 12-13.

Klein, P.F. Teaching styles. In H.E. Mitsel (Ed.), Encyclopedia of Educational Research. New York: The Free Press, 1982: 1927-1933.

Knapper, P.K. A decade review of college teaching research: 1970-1980. Canadian Psychology, 1981, 22: 129-145.

Kronk, A.K. & Shipka, T.A. Evaluation of faculty in higher education: A handbook for faculty leaders. Washington, D.C.: National Education Association, 1981.

Levinson-Rose, J. and Menges, R.J. Improving college teaching: A critical review of research. Review of Educational Research, 1981, 51, 403-434.

Lima, A.K. An economic model of teaching effectiveness. American Economic Review, 1981, 712: 1056-1059.

Medley, D.M. Teacher effectiveness. In H.E. Mitzell (Ed.), Encyclopedia of Educational Research. New York: The Free Press, 1982: 1894-1903.

Murray, H.G. A comprehensive plan for the evaluation of teaching at the University of Queensland. The University of Queensland Tertiary Education Institute, 1980.

Raths, J. Evaluation of teachers. In H.E. Mitzel (Ed.), Encyclopedia of educational research. New York: The Free Press, 1982: 611-618.

Rogers, W.T. Canadian educational research: Our responsibility. CSSE News, 1982, IX: 7, 8.1-8.7.

Scriven, M. The evaluation of college teaching. In Inservice. Syracuse, New York: National Council of States of Inservice Education, 1980.

Shavelson, R. and Dempsey-Atwood, N. Generalizability of measures of teaching behavior. Review of Educational Research, 1976, 46, 553-611.

Stapleton, J.J., Bulcock, J.W., Clifton, R.A. and Paschke, P.H. Teacher demand in Newfoundland and Labrador: The next ten years. St. John's: Institute for Educational Research and Development, Memorial University of Newfoundland, 1977.

Sullivan, A.M. and Skanes, G.R. Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 1974, 66: 584-590.

Williams, J. and Batten, M. The quality of school life. Hawthorn, Australia: The Australian Council for Educational Research, 1981.

ADDENDUM #1

The Dean's Memo

MEMORIAL UNIVERSITY OF NEWFOUNDLAND
St. John's, Newfoundland, Canada A1B 3X8

Faculty of Education
Office of the Dean

Telex: 016-4101
Telephone: (709) 737-8587/8

June 3, 1982

TO: Faculty Members, Faculty of Education

FROM: B. Paddock, Dean

SUBJECT: Committee on Faculty Evaluation.

For some time it has been my feeling and the feeling of a number of others in the Faculty of Education, that we should try to find a way to know and to record some information regarding the amount and quality of work which faculty members do both in scholarship and research and in teaching. Such information gathered regularly and sytematically could serve as a basis for annual, or more frequent, discussions between faculty members and Heads of their Departments and also as the basis for decisions regarding the granting of tenure, promotion and merit salary increases. In some instances, therefore, it would serve as information to be used in formative dialogue and in other instances it would be used in a summative sense.

I recognize that evaluation of faculty members work is a difficult and sensitive matter and, I suppose, some would contend, impossible. However, while we entertain such notions as improvement, merit salary increases, and promotion on the basis of the quality of one's work, we imply that evaluation is indeed possible and that judgements based on such evaluation are legitimate.

I have, therefore, appointed a committee to study the matter of evaluation of faculty members work and to make recommendations to the Dean regarding

1. Procedures appropriate for the evaluation of faculty members work

2. Criteria by which faculty members work should be evaluated.

The Committee is chaired by Dr. P.J. Warren and the other members are Dr. R. Kelleher and Dr. G. Sheppard.

...2/

The appointment of this committee does not imply that I am seeking ways to be negatively critical of faculty members work, nor does it imply that I am seeking ways of declaring faculty members incompetent or negligent. It is intended rather as a means of ascertaining if we can develop better ways of knowing the quantity and quality of a faculty member's work and if possible improving the work of the faculty.

I understand that in the near future the Committee will formally invite you to present any relevant ideas which you might have. In the meantime you should feel free to contact me or any member of the Committee if you have opinions which you wish to express regarding this matter.

I am confident that the work of the Committee will be viewed positively and conducted in a manner which will have a beneficial effect on all of us.

B.V. Paddock,
Dean.

/dp

ADDENDUM #2

Student Evaluation Subcommittee Terms of Reference

MEMORIAL UNIVERSITY OF NEWFOUNDLAND
St. John's, Newfoundland, Canada A1B 3X8

Department of Educational Administration

Telex: 016-4101
Telephone: (709) 737-7647/8

November 2, 1982

TO: Jeff Bulcock (Chairperson), Alan Griffiths, Terry Boak, and Marc Glassman

FROM: The Committee on Faculty Evaluation

SUBJECT: Preparation of a Position Paper on "Student Evaluation of Faculty Teaching Performance."

We are pleased that you have agreed to prepare a Position Paper for the Committee on "Student Evaluation of Faculty Teaching Performance." We would like for you:

(1) to summarize recent research findings regarding student ratings of faculty instruction;

(2) to suggest the purposes for which such student evaluation may be used, if at all, and the procedures that should be followed in their use; and,

(3) to examine available student rating forms with a view to recommending one or more (or any adaptation considered necessary) for possible use by our faculty. We understand that fulfilment of this final request may necessitate some piloting of available forms in selected classes at Memorial.

We request that you consider the advisability of adding a student to your group. It seems logical that, as student evaluations are intended for use by both students and faculty, both students and faculty members should participate in discussing and developing appropriate instruments. Randy Hawkins (an Education student on the Student Council with some responsibility for student evaluation), George French (a Faculty Council member), and Jeanette Blackwood (a member of our Faculty Council and active with the Education Society) have already indicated an interest in becoming involved in our work. Dr. Arthur Sullivan, who has done considerable research in this area, has also offered to help us in any way possible.

We have collected a number of readings concerning this matter and shall make these available to you shortly. In addition, we shall distribute a general bibliography on faculty evaluation for your use.

May we suggest the end of January, 1983, as a final date for the submission of your Paper?

We thank you most sincerely for your co-operation.

Yours truly,

P. J. Warren
Chairperson

ADDENDUM #3

A Respecified Model of Teaching Effectiveness

Prepared by

J.W. Bulcock

February 1983

ADDENDUM #3

A Respecified Model of Teacher Effectiveness

1. It was argued in the foregoing report that many models of teaching effectiveness are seriously misspecified. It was noted, for example:

(i) that different teachers have different teaching styles, yet many studies use global measures of teaching effectiveness, or they restrict their conception of style to dichotomous constructs (e.g., teacher centered vs. learner centered, or direct vs. indirect styles);

(ii) that few models are concerned with the dynamics of teacher effectiveness; therefore, most studies have been cross-sectional rather than longitudinal in design;

(iii) that there have been no studies in which efforts have been made to capture nonlinear relationships between the predictors and outcomes of teaching even though there are theoretical grounds for doing so;

(iv) that good teaching may be constrained by the time that teachers have to spend on other responsibilities such as research, writing and service, yet these constraints are seldom used as covariates in models of teacher effectiveness;

(v) that virtually all the research designs fail to take into consideration the fact that the typical model of teaching effectiveness is far from being orthogonal in view of the (sometimes) high correlations between predictor variables;

(vi) that in nonorthogonal (or collinear) models the redundancy of information provided by the independent variables leads to ambiguities when the researcher attempts to partition the variance of the dependent variable; in other words, when the multicollinearity is high the coefficients are unstable and difficult to interpret;

(vii) that the constructs in most models are primary sensory data or observables which are sometimes highly collinear, yet only modestly related to the outcomes of interest;

(viii) that few models consider the Jackson effect; namely, that the flow of influence between the class environment created by the teacher and the cognitive styles of the students is not one-way or recursive but two-way or reciprocal;

(ix) therefore, few models are designed to consider the importance of the classroom environment and the cognitive styles of students as sets of reciprocally interrelated intervening variables which mediate the impact of teacher styles on student achievements;

(x) that even in some of the better designed studies prior knowledge about their achievements can bias student ratings, thus accounting for a tautological component in the models of teacher effectiveness.

2. Sporadic suggestions were made when these limitations of the models of teaching effectiveness were discussed for coping more satisfactorily than at present with specification errors. Consider the following.

(i) To cope with the fact that different teachers have different styles of teaching a reconceptualization of the construct of "teaching style" may be in order. As a starting point the possibility of extending Broudy's (1972, 1974) typology could be considered. Such an exercise would invoke a two dimensional component of teaching style.

(ii) Dynamic models of teaching effectiveness are to be preferred over static models. This calls for longitudinal or time series designs rather than cross sectional ones.

(iii) The principle of generalizability should be invoked to override the principles of concreteness and inclusiveness on the grounds that the task of social modeling is not to mirror social reality, but rather to reduce it to a set of fundamental and manageable features. This calls for structural equation modeling; that is, for the analysis of the covariance structure of complex sets of relationships. In structural equation modeling the analyst uses multiple indicators of the constructs to be used. Confirmatory or explanatory factor analysis or principal component analysis are methods commonly used to compute linear combinations of these indicators. These linear combinations are likely to be collinearity reduced models, though not collinearity free ones.

(iv) If the structural equation model is highly collinear, ridge regression could be considered as an alternative to ordinary least squares regression or full information maximum likelihood (FIML) estimation.

(v) To cope with feedback estimation two stage least squares or simultaneous equation estimation of the FIML variety could be considered. Multiple feedback (i.e., where $N > 2$ feedback loops have to be estimated) cannot usually be coped with by 2SLS or FIML estimation methods, so two-stage normalization ridge regression might have to be considered.

(vi) To cope with the bias in student ratings their cognitive styles should be used as covariates.

The model depicted in Figure A3.1 takes these specification problems into account. However, Figure A3.1 is composed of structural elements which have to be disaggregated if the model is to be estimated.

Figure A3.1 About Here

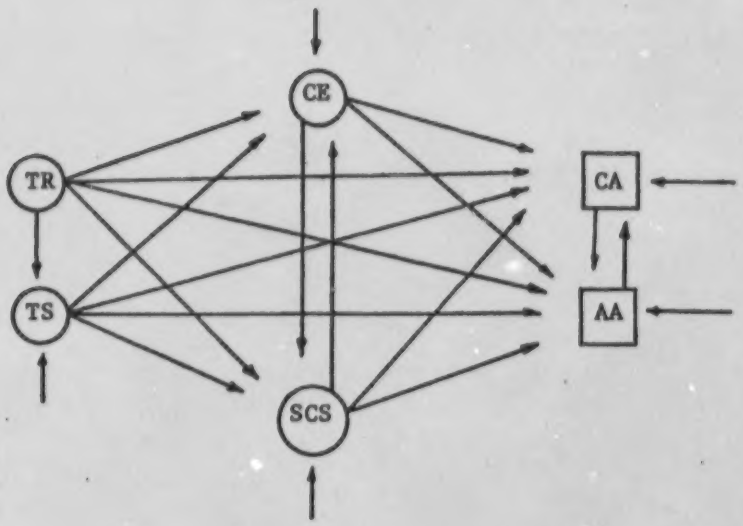
TR-Teacher Resources. This is conceived of as a reflective construct with three dimensions: a resource dimension, a personality dimension and a perception of teaching dimension. These are speculative notions, however, which means that exploratory factor analysis is in order. Some tentative indicators of each dimension might be as follows.

Resource dimension - qualifications, experience, age, sex, research and publications, professional society activities, service on graduate student committees.

Characteristics or personality dimension - humour, enthusiasm, communication skill, helpfulness, consideration toward students, open mindedness, stimulating, informativeness, knowledgeable, example, friendliness, approachability.

Perception of Teaching - the relative importance of teaching role; the degree of faculty member responsibility for (a) the students' learning, (b) student advising, (c) being accessible to students, (d) improving instructional effectiveness,

Figure A3.1. A Conceptual Model of Teaching Effectiveness



Key: TR = teacher resources, TS = teaching style, CE = class environment, SCS = student's cognitive style, CA = cognitive achievement, AA = affective achievement.

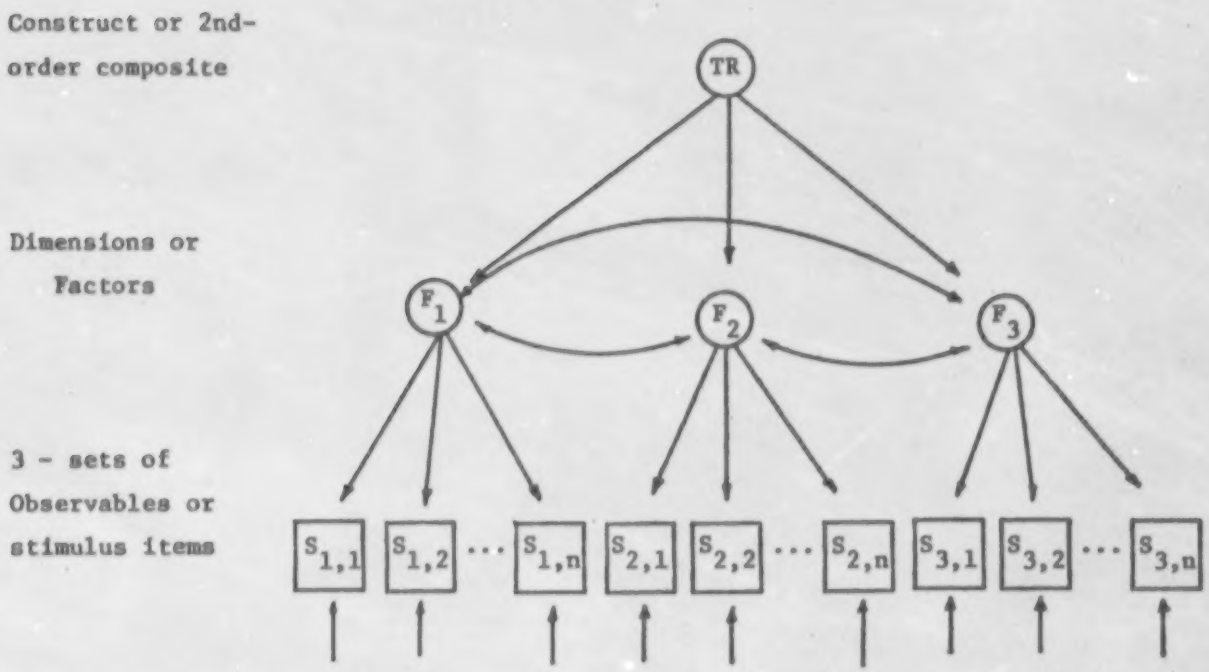
(e) helping students who have difficulty understanding the material, (f) keeping lectures up to date, (g) students' progress in the course, (h) making the course relevant to real life situations, (i) preparing students to cope with more advanced subject matter in later courses, (j) developing students' sense of personal responsibility (confidence, self-discipline), (k) the amount of reading a student does in a course, (l) the amount of time a student spends on a course.

A factor model of Teacher Resources, then, would consist of a second order composite. The model is illustrated in Figure A3.2.

Figure A3.2 About Here

TS - Teaching Style. Though there are numerous pronouncements on teaching styles there is little theory that the analyst can draw on. See Kleine (1982) for details. Here we construct our own theory by assuming that the two most important dimensions of teaching style are (a) orientation to knowledge and (b) orientation to students. Harry Broudy notes that the teacher may prefer either to emphasize the achievement of cognitive goals or to emphasize the process whereby the goals may be achieved. The first emphasis he referred to as didactic teaching; the second as heuristic teaching. The didactic style is a type of teaching which

Figure A3.2. A Second-order Factor Composite of the Teacher Resources Construct



where $S_{1,1}$ thru $S_{1,n}$ = resource dimension items, $S_{2,1}$ thru $S_{2,n}$ = personality resources dimension items, $S_{3,1}$ thru $S_{3,n}$ = perception of teaching dimension items, F_1 = resource dimension, F_2 = personality dimension, F_3 = perception of teaching dimension, TR = a second order reflective composite - teaching resources. The curved relationships are the interfactor correlations. The weights on the single headed paths are factor loadings which are used to weigh the scores in constructing linear combinations of variables. Variables in rectangles are observables whereas circled variables are unobserved or latent constructs.

tries to impart the skills, the attitudes and the information appropriate to the knowledge goal. The heuristic approach places most emphasis on the procedures involved in the act of discovery, in the art of solving problems; that is, the logic of inquiry.

Though Broudy recognizes that the teacher's orientation to students is important he only deals with one end of the continuum which he labeled "philetics". The word "philetics" is derived from the Greek word "philos" meaning "love of". Philetic teaching focuses on the improvement of relationships between teacher and pupil, the pupil and other pupils, and the pupil and himself. But at the other extreme is the impersonal teacher who is ambivalent or neutral insofar as teacher-student relationships are concerned.

A division of the two dimensions of teaching gives the four fold typology depicted in chart A3.1. The four types or styles of teaching are labelled "progressive-heuristic", "philetic-didactic", "traditional-heuristic" and "formal-didactic" in cells I through IV respectively.

Chart A3.1 About Here

Teaching style may be conceived of as a reflective construct. Since its empirical status is as yet ambiguous exploratory factor analysis is called for. A model which identified Broudy's three modes of teaching may be depicted as in Figure A3.3. Characteristic items might be as follows.

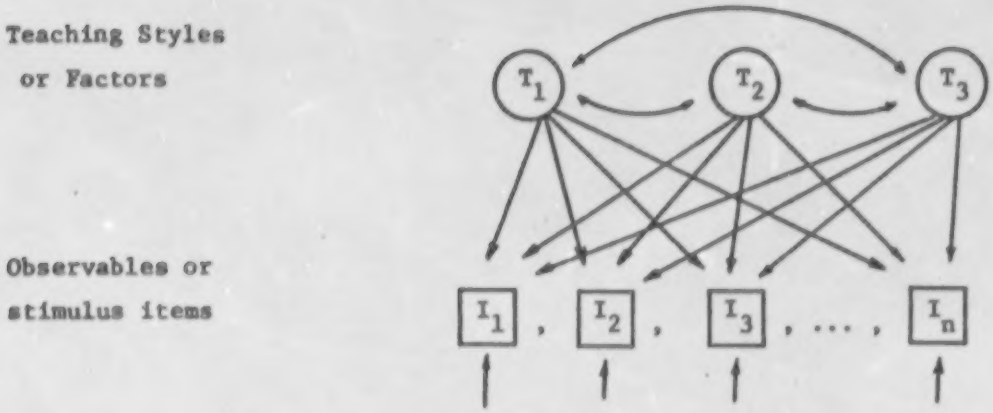
Figure A3.3 About Here

CHART A3.1

A TYPOLOGY OF TEACHING STYLES

Orientation to Students	Orientation to Knowledge: focus on means	Orientation to Knowledge: focus on ends
Concerned with improved relationships	I Progressive - heuristic learning through doing	II Philetic - didactic learner-centered indirect instruction
Neutral orientation toward student-teacher relations	III Traditional - heuristic	IV Formal - didactic teacher-centered direct instruction

Figure A3.3. Factor Model of Teaching Styles



I, I_2 thru I_n = stimulus items, T_1 = didactic orientation, T_2 = heuristic orientation, T_3 = philetic orientation. Variables in rectangles are observables while those in circles are latent constructs.

Orientation to Knowledge Items

The instructor:

expects students to know how to learn
gives challenging problems
covers the course material thoroughly
is familiar with the latest research studies

Evaluate your progress in this course in terms of:

knowledge of the course content
knowledge of the major principles and theories
problem solving ability
how professionals in this field gain new knowledge
the development of your creative capacity

Orientation to Students

The instructor:

establishes good rapport with students
genuinely concerned with student progress
always speaks positively about students
greets his students outside the classroom
helps students when they have trouble understanding the material
has a lively sense of humor
speaks at a rate which allows time to take notes

CE - Classroom Environment. On the basis of the relationships depicted in Figure A3.1 it is hypothesized that the classroom environment will be responsive to both the teaching resources of the instructor as well as to his teaching style. The cognitive styles of the students will also influence the classroom environment as well as being influenced by it. Here we invoke a social learning theory

characterization of the classroom environment (Bandura, 1977). In terms of social learning theory teachers hold expectations for their students' achievements and scholastic performances; they provide opportunities for their students to practice and learn these behaviours; they reinforce their students' performances and provide models as a means whereby these behaviours may be learned.

Encouragement Items

encourages the pursuit of excellence
encourages student participation in class activities
encourages independent study and outside reading
encourages students to respond to one another in class
encourages student initiative
sets realistic expectations

Opportunity to Learn Items

The instructor:

identifies his instructional objectives
makes effective use of a text
makes maximum use of the available class time
makes provision for extra classes or tutorial assistance
gives too many examinations
gives too many assignments
prescribes outside reading
prescribes non-reading assignments
makes effective use of media resources
makes effective use of examples or illustrative materials

Reinforcement Items

The instructor:

makes helpful comments on assignments
makes helpful comments on term papers or examinations
is niggardly in his allocation of marks
returns papers promptly
receives student questions politely

Role Model Items

The instructor:

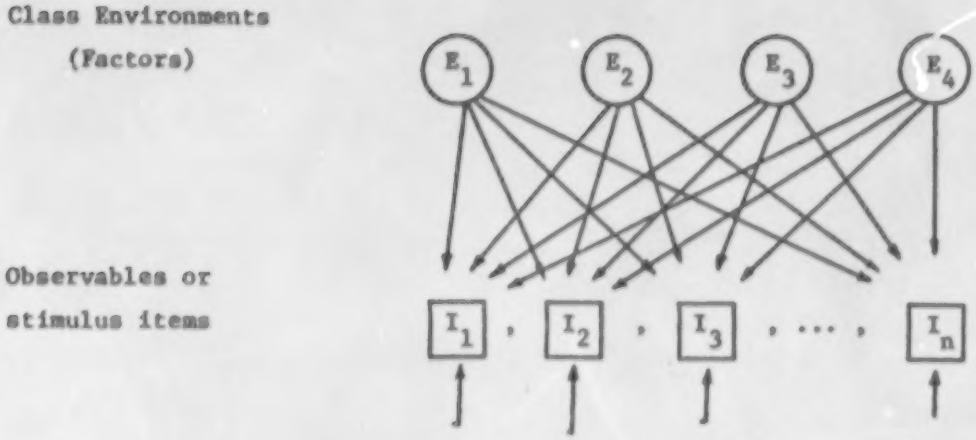
keeps up with the latest literature
is an active participant in learned society meetings
is a well-known campus figure
in lectures refers to recent books and scholarly articles
publishes articles
conducts research
is egotistical
is a person I admire

A factor model designed to capture these dimensions of he classroom environment is depicted in Figure A3.4.

Figure A3.4 About Here

SCS - Student's Cognitive Style. The notion of cognitive style is dominated by a Piagetian developmental orientation - the four (sometimes five) stages of thinking are well known. In recent years, however, cognitive psychologists have formulated an information processing orientation to the question of cognitive style. Information processing proceeds in three stages: a cueing strategy stage, a discourse strategy stage and an inferencing stage.

Figure A3.4. A Classroom Environment Model



Key: I_1 thru I_n = stimulus items, E_1 = Encouragement Structure
E_2 = Opportunity to Learn Structure, E_3 = Reinforcement Structure
E_4 = Role Model Structure.

Cueing strategies are used when information is being decoded - especially written (graphic) information. A bottom-up strategy is dependent on text based cues which are predominantly graphic and phonic. A top-down strategy is dependent on one's intuitive grasp of language rules; namely, syntactic and semantic cues. At the college level bottom-up strategies may be constants; hence, of little interest.

Discourse processing strategies are used to organize decoded information in order to facilitate encoding. At the most rudimentary level information is concatenated or strung together. Higher level processing involves relating two pieces of information. At the highest level of discourse processing one may refer to a synthesizing strategy. Here the decoder may demonstrate mastery of such syntactic devices as causal connectives, temporal conjunctions, referential story pronouns and referential formal repetition. These are grammatical usages which are essential for the integration of ideas both causally and sequentially.

The third stage in information processing involves the use of hypothesizing, predicting and inferencing. After new information has been organized it must be linked to what the individual already knows. At the most mundane level the inferences will be of an explicit or informational nature. Characteristic inferences will be of the temporal and spacial variety and for the most part will consist of adding details to what is already known. They may be referred to as informational inferencing strategies. At a higher level

causal or interpretive inferences will be used. These may be called enabling strategies. At the highest level of inferencing evaluative and prescriptive inferences are called for which involve the prediction or anticipation of outcomes and the making of value judgments. These are called evaluative strategies.

Different styles will utilize different information processing strategies. A model of information processing, then, will be composed of three strategies; namely, a syntactic-semantic strategy, a discourse strategy and an inferencing strategy. They may be depicted as in Figure A3.5.

Figure A3.5 About Here

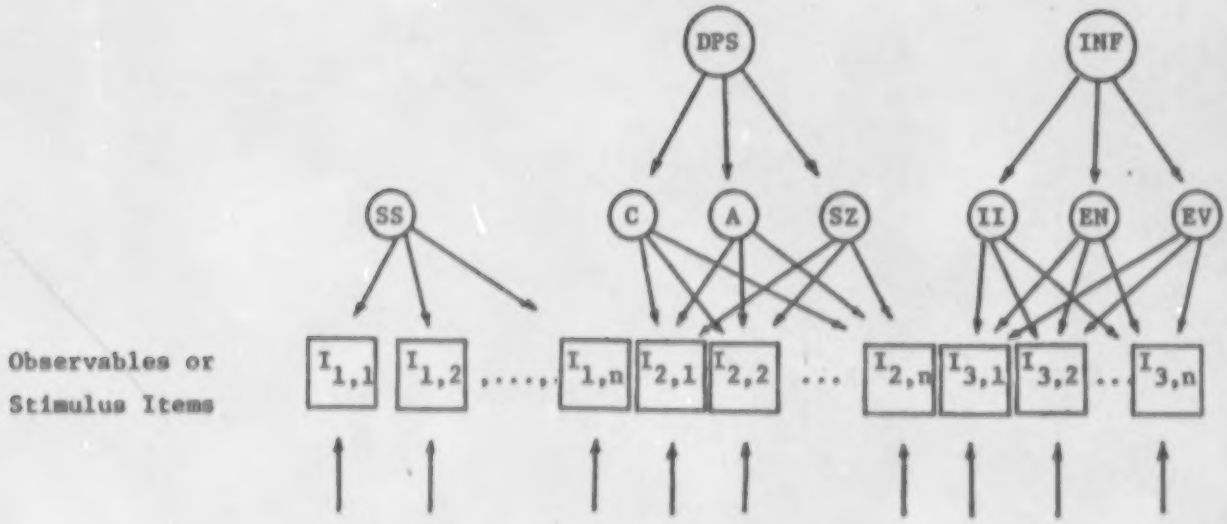
CA - Cognitive Achievement. Consider, for example, achievements in two distinct forms of thought such as English Literature and Mathematics.

AA - Affective Achievement. Consider two conative factors such as motivation and academic expectations.

Consider, too, the notion that achievement stimulates achievement so that cognitive and affective achievements are reciprocal. A multiple feedback learning system model is depicted in Figure A3.6.

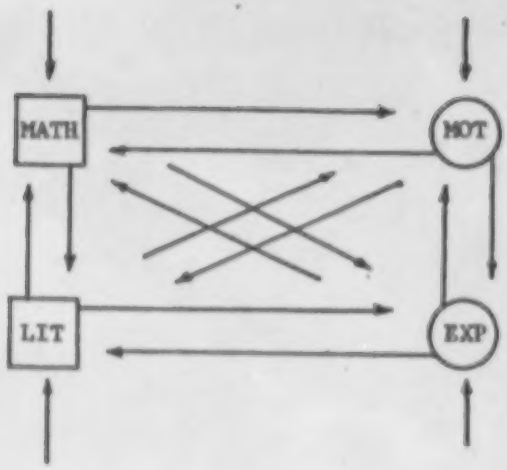
Figure A3.6 About Here

Figure A3.5. Factor Model of Student's Cognitive Styles



Key: Where $I_{1,1}$ thru $I_{1,n}$ = syntactic-semantic stimulus items, $I_{2,1}$ thru $I_{2,n}$ = discourse processing strategies stimulus items, $I_{3,1}$ thru $I_{3,n}$ = inferencing strategies stimulus items, SS = syntactic-semantic strategy, C = concatenating strategy, A = associating strategy, SZ = synthesizing strategy, II = information inferencing strategy, EN = enabling inferencing strategy, EV = evaluative inferencing strategy, DPS = discourse processing strategy, INF = inferencing strategy.

Figure A3.6. A Multiple Feedback Learning System Model



Key: MATH = mathematics, LIT = English literature
MOT = motivation, EXP = academic expectations

A disaggregated structural equation model of teacher effectiveness incorporating the constructs presented in Figures A3.2 thru A3.6 is depicted in Figure A3.7.

Figure A3.7 About Here

Model Equations

(1) $T_i = f(x_1, x_2, \ldots, x_k)$

Model #1. Teaching styles model

where T = teaching styles

x_i = the amount of teaching resource i chosen by the individual to contribute to teaching style

(2) $E_i = f(x_1, \ldots x_k, t_1, t_2, \ldots, t_m, c_1, c_2, \ldots, c_p)$

Model #2. Classroom environment model

where E = classroom environment

x_i = amount of teaching resource i contributing to the classroom environment

t_i = the effect of teaching style i on the dimensions of the classroom environment

c_i = the effect of student cognitive style i or information processing strategy i on the classroom environment

(3) $C_i = f(x_1, \ldots, x_k, t_1, \ldots, t_m, e_1, e_2, \ldots, e_n)$

Model #3. Student cognitive styles model

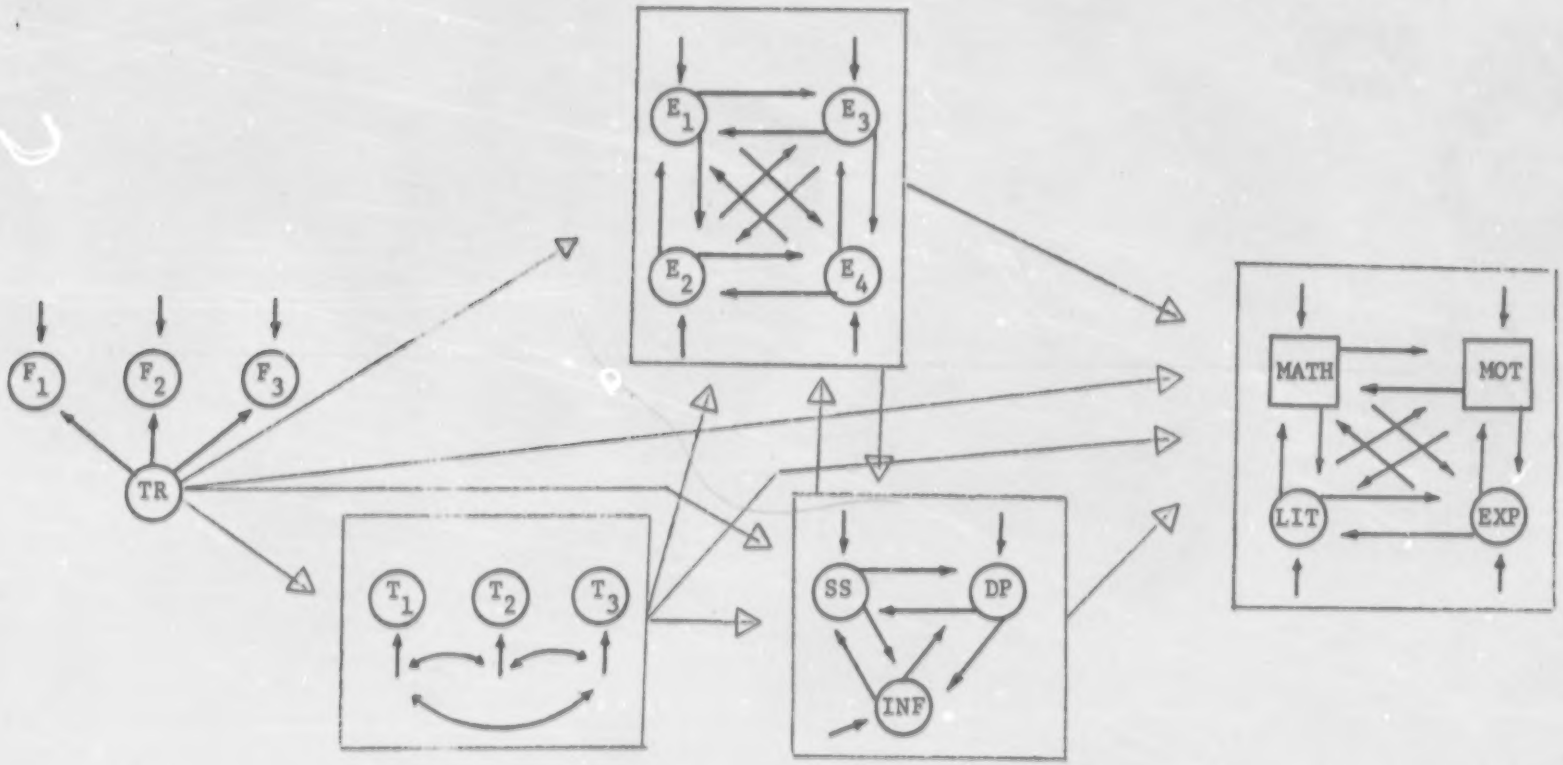
where C = student cognitive style

x_i = amount of teaching resource i contributing to the student's cognitive style

t_i = the effect of teaching style i on the student's cognitive style

e_i = the effect of classroom environment dimension i on the student's cognitive style

Table A3.7. A Disaggregated Structural Equation Model of Teacher Effectiveness



Key: TR = teaching resources, F_1 = resource dimension, F_2 = personality dimension, F_3 = perception of teaching dimension, T_1 = didactic teaching orientation, T_2 = heuristic reaching orientation, T_3 = philetic teaching orientation, E_1 = classroom expectation structure, E_2 = opportunity to learn structure, E_3 = classroom reinforcement structure, E_4 = teacher role model, SS = syntactic-semantic information processing strategy, DP = discourse processing strategy, INF = inferencing strategy, MATH = mathematics achievement, LIT = achievement in English Literature, MOT = motivation, EXP = academic expectations.

(4) $A_i = f(x_i, \ldots x_k, t_1, \ldots t_m, e_1, \ldots e_n, c_1, \ldots c_p)$

Model #4. Student achievement model

where A = the student's cognitive and affective achievement outcomes

x_i = the amount of teaching resource i contributing to the student's cognitive and affective achievement

t_i = the effect of teaching style i on the student's cognitive and affective achievement

e_i = the effect of classroom environment dimension i on the student's cognitive and affective achievement

c_i = the effect of student cognitive style i on his cognitive and affective achievement

(Addendum 3)

References

Bandura, A. Social Learning Theory. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Broudy, H.S. Didactics, heuristics and philetics. Educational Theory, 1972, 22, 151-161.

Broudy, H.S. Three modes of teaching and their evaluation. Mimeograph. Memorial University, Department of Educational Foundations, 1974.

Kleine, P.F. Teaching styles. In H.E. Mitzell (Ed.) Encyclopedia of Educational Research. New York: The Free Press, 1982: 1927-1933.

ADDENDUM #4

A Proposed Compromise Solution to the Problem of Evaluating Teacher Effectiveness *

A Position paper prepared by

J.W. Bulcock

* The proposed compromise solution was discussed by members of the Committee on Student Evaluation of Faculty Teaching Performance. A consensus could not be reached largely because to some members it seemed as if the compromise was merely an attempt to bring in summative evaluation through the back door. Thus, the proposed compromise is presented as an addendum to the report for the information of interested Faculty members.

A Proposed Compromise Solution to the Problem of Evaluating Teacher Effectiveness

On the one hand summative evaluation is here to stay, on the other the summative model of teacher effectiveness is seriously misspecified. On the one hand handbooks on the terms and conditions of faculty employment make the evaluation of a faculty member's teaching performance manditory, on the other hand administrators are reluctant to conduct such evaluations; indeed, would far sooner depend on student ratings. On the one hand the summative evaluation of a faculty member's teaching performance minimizes arbitrary personnel decision-making and counters unprofessional judgments, on the other hand its consequences through bureaucratizing the procedure may easily become counter productive and self-serving.

These dilemmas are another reason why summative evaluation methods are unpopular and controversial. What can be done? One is reminded that a good deal of information about a faculty member's teaching is available without resorting to student, administrator or colleague ratings of teaching, or any other summative approach. Teaching in a broad sense includes numerous activities. Some are constants such as:

the provision of laboratory instruction,

the preparation, administration and grading of tests and examinations,

the reporting of course grades.

Other activities are optional such as:

the provision of seminars for small groups of students,

the provision of student conferences,

the counselling or advising of students,

the supervision of undergraduate student research,

the supervision of graduate student research, and

the provision of reading courses.

Some activities are a consequence of teaching such as:

the writing of letters of recommendation,

the submission of library requests for book purchase,

the submission of requests for AVA materials,

the submission of requests for laboratory materials and

the preparation of teaching aids such as models, charts, diagrams, graphic and photo-graphic displays, etc.

Effective teaching can generate a chain reaction such as:

the presentation of departmental or faculty seminars,

the presentation of work shops,

the presentation of invitational addresses to service clubs, professional organizations and conferences,

the presentation of invitational lectures at other universities,

the presentation of inspirational addresses at speech days, commencement exercises, graduation ceremonies and the like.

And a host of miscellaneous activities may also be identified as an integral part of the teaching function:

the preparation of new lectures,
the preparation of a new course,
the coordination of multi-sectioned courses,
acting as rapporteur at a conference,
the presentation of papers at a World Congress,
the presentation of papers at refereed conferences,
the presentation of a paper at a symposium,
the presentation of papers at learned society meetings,
student teaching supervision,
after dinner speaking,
chairing panel discussions,
speaking as the member of a panel, and
competing for a teaching reward.

These teaching functions are not exhaustive. But they do indicate the wide range of activities which may be regarded in a collective sense as a proxy for effective teaching. One would be reluctant to rate a faculty member who has presented fifty or so papers at learned society conferences and who has been an invitational speaker at numerous conferences as an ineffective teacher. It is advisable for faculty members who are eligible for tenure or promotion to maintain an annual record of their teaching roles and responsibilities, and when necessary to communicate the information to Departmental Heads and tenure committees.

ADDENDUM #5

Why is the Matter of Faculty Evaluation a Pressing Issue in the 1980s?*

A Position Paper Prepared by

J.W. Bulcock

January 1983

---

* Though the Committee on Student Evaluation of Faculty Teaching Performance discussed the reasons for the salience of the topic, it was noted that it falls outside the Committee's terms of reference. This position paper on the demographic basis of the problem is presented, therefore, as an addendum to the report for the information of interested Faculty members.

The Mid-Career Promotion Squeeze

Why is the matter of faculty evaluation a pressing issue in the 1980s? What differs today from yesterday? An economic argument would be persuasive except that decline management, retrenchment and constraint in the Faculty has been a way of life for almost a decade. For example, the Parsons' Committee on declining enrolments was struck in 1973 and reported in 1974-1975. The most persuasive argument is probably a demographic one.

Demographic issues are slow to develop. They are diffuse and long lived. In general, population changes are glacier-like. Though slow to proceed and massive in scale,

they mount in intensity. In market economies as the pressure of change increases the impact of demographic forces tends to be ignored until the last moment. In fact, demographic phenomena are usually not responded to until they are experienced in economic terms; for example, when new schools have to be built or old schools closed and sold. That underlying these market place decisions there exists pervasive demographic forces may only dimly penetrate human consciousness.

As early as 1965, Alan Cartter in a paper presented at the Proceedings of the American Statistical Association predicted an impending oversupply of teachers and university professors. Nevertheless, teachers and professors were trained in increasing numbers through the late 1960s and early 1970s. The problem stems from two demographic anomalies within a single generation (thirty years). First, there was a baby boom due to high birth rates in the 1950s and early 1960s which was followed virtually without warning by a baby bust - steeply declining birth rates in the 1970s to the present. For every hundred children under five in 1965 today there are only sixty-five.

In the baby boom years there was a need for more classrooms, more teachers and eventually more professors. Today, the system is operating in reverse. Many of the once overcrowded classrooms lie empty. The persistent teacher surplus over the past several years is aggravated by the increasing number of reserve teachers - those women who graduated in the 1960s and whose family responsibilities have declined - seeking reentry into teaching.

The Promotion Squeeze for Teachers

By the mid-1980s the large group of teachers produced in the mid-1960s and early 1970s will be at mid-career -- roughly 35 to 44. They are established in their careers. They can foresee in broad outline what the next 20 to 25 years will bring in terms of advancement and career satisfaction. For many the prospects look dim.

Several factors are at work. As the baby boom cohort matures the number of persons in the 35-44 age range will swell. Between 1977 and 1987 there will occur a 45 percent increase in the number of 35-44 year-olds. Three effects are predictable: a life-cycle effect, a generational effect and an economic effect. For the 35-44 year-old cohort life-cycle effects are important. Typically job dissatisfaction for professional workers increases up to age 35, falls off, and then increases again at mid-life, around 50. The dissatisfaction, however, can be offset by career advance - increased occupational responsibility and concomitant salary increase, promotion, increased autonomy in the sphere of work, assumption of leadership roles and the like.

There are also generational effects. Each cohort of workers tends to be better educated; hence, tends to set its sights higher in terms of what it regards as satisfying. Here, working conditions, salaries, working hours, job status and a host of intangible, situation specific, considerations may operate. Economic effects should not be overlooked. Job dissatisfaction may fluctuate in tandem with the overall state

of the economy and with the worker's sense of job security. Well qualified teachers with permanent appointments may be more satisfied in periods of economic recession than in periods of rapid economic growth. Thus, at this time of recession job dissatisfaction due to life-cycle and generational effects may be significantly offset by an economic effect. This may be especially the case with women teachers since in times of recession the employment prospects for women are more adversely affected than they are for men.

Despite the countervailing influence of economic factors the imbalance in the age structure of today's teaching profession may be the most critical factor. When the well-educated baby boomers try to move past the lower rungs of the occupational ladder the competition for status will intensify. The rules of the promotion game will operate in zero-sum fashion.

The contradiction is this. From 1955 to about 1970 the educational labour market was elastic enough to absorb the ever increasing supply of highly qualified teachers and professors. In these years every youngster had the opportunity and the responsibility not just to graduate from high school but also from technical college or university. Efforts to turn out teachers, technicians and scientists was justified on the grounds that young Newfoundlanders had an obligation to fill positions that were opening up in a rapidly changing society.

But then the Catch 22 situation arises. Four years in the Faculty of Education was followed by graduation and job

market entry as a school teacher. Compared to earlier generations of teachers these baby boomers were the best qualified, the most articulate and probably as committed as any in history. But in recent years these graduates have been informed that the rules have been changed. Their career advancement is dependent on their acquisition of advanced credentials. A master's degree is the minimum qualification for promotion. And as the mid-career promotion squeeze intensifies the master's degree is unlikely to prove sufficient. To face tomorrow's competition the Ph.D. is becoming necessary. Ph.Ds., however, are only offered by mainland or foreign universities at a cost which is beyond the means of most married teachers.

Given the growing bulge of 35-44 year-old teachers the educational system in Newfoundland cannot produce the number of senior appointments in education that provide the status that many ambitious teachers strive for. The problem is one besetting all Provinces of Canada to varying degrees. No matter what educational level is required for entry into these jobs today's teacher force is capable of achieving that level. In Malthusian terms the educational level of the teacher force is growing geometrically while the number of jobs that require advanced levels of education has grown arithmetically. Thus, there is a marked disjunction between what the educational system expects and promises and what it, in fact, can offer. The contradiction is manifest not only in terms of a current teacher surplus but also in terms of a

career squeeze for many highly qualified and ambitious teachers now at mid career. The attendant frustration may have deleterious effects on teacher morale, professional commitment, job satisfaction and performance as many teachers conclude at mid-career that they are locked into a static occupational pattern with few prospects of meaningful career advancement. It must be obvious that one consequence of the mid-career promotion squeeze will be manifest in changing attitudes toward the teacher education program and especially the graduate program at Memorial University.

The Promotion Squeeze for Professors

Though the demographic forces affecting the mid-career promotion squeeze have been discussed with reference to school teachers the promotion squeeze for professors in Canadian universities is the same in kind. In relation to other ranks disproportionately large numbers of Associate Professors are vying with one another for promotion. Unlike the school teacher who in theory can pursue more advanced credentials, the Associate Professor has long since reached the end of the credential road. Over ninety percent have an earned doctorate.

The Catch 22 situation for the Associate Professor is much the same as for the school teacher. At the crux of the problem is the official rhetoric which in effect states that promotion to the rank of Professor is reserved for faculty members of outstanding merit. Merit, that is, in terms of scholarship, teaching effectiveness, professional service, and

contribution to the University and the community. This is normal curve rhetoric. In any group large enough to allow the law of large numbers to operate (e.g., N > 24) a fraction will fall under the tail of the distribution justifying the label "outstanding". The snag is that the properties of the normal curve capture a mathematical, not a psychological reality.

The distribution of achievement in tensor algebra of a hypothetical class of Nobel laureates in physics would be no less "normal" in the statistical sense than that of a grade six class in arithmetical problem solving. Only a handful of the Nobel laureates would fall under the "outstanding" tail of the distribution and, by the same token, only a few of the grade sixers would qualify as "outstanding". In the same way a few of the Associate Professors at Memorial could be labeled "outstanding" in terms of scholarship. A few could be labeled outstanding in terms of teaching, and so on. Each criterion - scholarship, teaching, etc. - could be weighted in order that an overall aggregate faculty member performance score could be computed. Or could it?

What can be summarized about the promotion squeeze for professors? We have noted the imbalance in the age structure of the professoriate which in view of the normal curve rhetoric is accounting for a promotion squeeze. The competition is becoming increasingly zero-sum. As more Associate Professors vie for promotion the promotion pressure intensifies. As more junior faculty move into Associate Professor rank the more

congested this rung of the career ladder becomes.

The issue is aggravated by the problem of the double standard. The standards prevailing in today's promotion mill are different from those operating in yesterday's mill. Professors appointed in the 1950s are close to retirement. They served during the expansionary period in post-secondary education, the 1960s. In this growth management era the Ph.D. was a safe route to the Associate Professorship. It was not uncommon for recent Ph.Ds. in education with some practical experience in the school system to be given a first appointment at the Associate Professor level. The department headship was a recognized route for promotion to Professor rank. It was tempting in the expansionary era for the relatively inexperienced academic to begin an administrative career prior to establishing a solid scholarly record. Thus, it is not uncommon today for junior faculty to be assessed for tenure and promotion by senior faculty who have less distinguished scholarly records than their junior colleagues. Similarly, because of the promotion squeeze, Associate Professors with superior records and resumes than their senior colleagues may be turned down for promotion because they are competing with peers with even more distinguished records. It is these kinds of anomalies which foster the frustration, disappointment and disilusionment of those stuck in the promotion mill. The inherent conflicts attributable in part to the promotion pressure tends to erode peer relations and to promote factionalism.

Oddly enough the problem was anticipated at Memorial University over six years ago. The Stapleton Report of 1977 (p. 13) on teacher demand noted that:

> Education in the noncompulsory sector will become a declining rather than a growing industry necessitating concomitant changes in management styles. Such changes will be accompanied by conflict between vested in-group interests, including a mid-career promotion squeeze ... Such conflicts will be accompanied by new ideological emphases, and probably a decline in morale.